稳健的对数最优策略理论研究与实践

投资与赌博的共同点是资产收益的随机性。通过概率分布函数刻画随机性，在结合运筹学的知识就可以构建资产配置的模型。模型所优化的指标不同，对应的长期资本积累的表现也不一样。怎样确定优化指标使得长期资本积累最多以及如何稳健的实现这一目标具有重要的现实意义，这也是本文即将重点分析的两个问题。

1.1 问题引入：经典赌博资金分配问题

假设有一个赌徒面对一个赌局，使用一部分资金押注。若为胜局，押注资金翻倍，败局时，押注资金全部失去，胜局的概率记为$\mathrm{p} \in [0.5,1)$。该赌徒进行连续的赌博，每一次的规则和胜局概率保持相同，使用怎样的策略才能使该赌徒赢得更多的资金。

分析：我们假设理想的条件，即没有最小下注限制，下注资金可以为实数域任意数字。赌徒不能改变赌局的规则和概率，只能调整自己每轮下注的数量。若赌徒每次使用全部资金进行赌博，则他输光的概率为$\lim_{\mathrm{n}\to\infty}(1-p^n)$=1.若赌徒不下注，资金得不到正向的增长。保证不输光以及获得最优的资本增长的要求下，最优的下注比例一定在[0,1)区间内，因为每轮赌局的规则概率都相同，每轮最优的下注比例必然相同。

记初始情况，赌徒资产为X_0，每次使用$\mathrm{b} \in [0,1)$比例的资金下注。在第 n 局后，赢了 S 局，输了 n-S 局。n 局后的资金为$\mathrm{X}_\mathrm{n} = X_0(1+\mathrm{b})^S(1-\mathrm{b})^{n-S}$。记

$$\mathrm{f}_\mathrm{n}(\mathrm{b}) = \log\left(\frac{X_n}{X_0}\right)^{\frac{1}{n}} = \frac{s}{n}\log(1+b) + \frac{n-s}{n}\log(1-b)$$

$\mathrm{f}_\mathrm{n}(\mathrm{b})$度量了资金的指数增长速度。它的期望为：

$$\mathrm{f}(b) = E\left[\log\left(\frac{X_n}{X_0}\right)^{\frac{1}{n}}\right] = E\left[\frac{s}{n}\log(1+b) + \frac{n-s}{n}\log(1-b)\right]$$
$$= p\log(1+b) + (1-p)\log(1-b)$$

为了使期望最大，令导数为 0：

$$\mathrm{f}'(\mathrm{b}) = \frac{\mathrm{p}}{1+\mathrm{b}} - \frac{1-\mathrm{p}}{1-\mathrm{b}} = \frac{2\mathrm{p}-1-\mathrm{b}}{(1+\mathrm{b})(1-\mathrm{b})} = 0$$

解得$\mathrm{b} = 2\mathrm{p}-1 \in [0,1)$.此时$\mathrm{f}''(\mathrm{b}) = \frac{-\mathrm{p}}{(1+\mathrm{b})^2} - \frac{1-p}{(1-b)^2} < 0$。则当$\mathrm{b} = 2\mathrm{p}-1$时，$\mathrm{f}(w)$取(0,1)上的极大值，即达到最大的期望增长速度。

此外，若没有$p \in [0.5,1)$的条件，当 p<0.5 时，最优 b 为负，即以 b 比例的资金“看空”这场赌博可以获得资金增长，这在金融市场中即对应着卖空操作，在简单的赌局中没有卖空选项，最大的增长率在 b=0 处达到。当 p=1 时，对应着必胜的赌局，“全压”是最优的增长策略。

此问题和解答揭示出，在规则概率不变的赌局中，存在唯一的固定的最优资金投入比例，使得资本的期望增长速度最大。

1.2 拓展至多元独立同分布的情况

对一元离散的经典赌博情况进行拓展，考虑多元的连续收益分布情形。举个例子，现实生活中的股票市场收益是连续多元的。假设资产价格比（定义为下一时刻价格与上一时刻价格之比）向量$\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{n})^\mathrm{T}$服从分布F（$\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n}$），资产配置向量为$\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$。

此情况下，最大化资本期望增长速度转化为一个优化模型：

$$\mathbf{b} = \underset{\boldsymbol{b} \in B}{\operatorname{argmax}}\, E\log(\mathbf{b}^\mathbf{T}\mathbf{X}) = \underset{\boldsymbol{b} \in B}{\operatorname{argmax}} \int \log(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X})\, d\mathrm{F}(\mathbf{X})$$

其中 B 表示可行域，体现出$\mathbf{b}$的约束条件。

资产价格比向量分布是独立同分布的，使资本期望增长速度最大的配置也是固定不变的。使资本期望增长速度最大的策略定义为**对数最优策略**。

2 考虑信息的影响，非独立同分布情况下的策略

实际上，市场上的一些信息对资产的未来收益是有预测作用的，这也直接影响着资产配置。将信息的影响引入投资时，连续投资环境下，资产收益的独立同分布假设变得不合适。不独立同分布条件下，资产增长策略会有什么变化？解答这个问题之前，先考察每一轮投资时，信息的引入对最优资本增长速度的期望的影响。

在某轮投资时，记信息为$\mathrm{Y} = \mathrm{y}$,给定信息后，资产价格条件分布为$\mathrm{F}(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n} | \mathrm{Y} = \mathrm{y})$，此时最优增长速度的配置记为$\mathbf{b}^* = \underset{\boldsymbol{b} \in B}{\operatorname{argmax}} \int \log(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X}) d\mathrm{F}(\mathbf{X} | \mathrm{Y} = \mathrm{y})$，增长速度的期望为$\int \log(\boldsymbol{b}^{*\boldsymbol{T}}\boldsymbol{X}) d\mathrm{F}(\mathbf{X} | \mathrm{Y} = \mathrm{y})$.
没有此信息时最优的配置记为$\mathbf{b_0} = \underset{\boldsymbol{b} \in B}{\operatorname{argmax}} \int \log(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X}) d\mathrm{F}(\mathbf{X})$，增长速度的期望为

$\int \log(b_0^{\boldsymbol{T}}\boldsymbol{X})d\mathrm{F}(\mathbf{X}|\mathrm{Y}=\mathrm{y})$。故因为信息的引入，增长速度期望有：

$\Delta V_{Y=y}=\int \log(\boldsymbol{b}^{*\boldsymbol{T}}\boldsymbol{X})d\mathrm{F}(\mathbf{X}|\mathrm{Y}=\mathrm{y})-\int \log(\mathbf{b}_{\mathbf{0}}^{\boldsymbol{T}}\boldsymbol{X})d\mathrm{F}(\mathbf{X}|\mathrm{Y}=\mathrm{y})$的增长。由$b^*$的定义知$\Delta V_{Y=y}\geq 0$，这说明理论上信息Y的引入不会降低资本最优增长速度，不同信息带来的速度增加会有怎样的不同？这需要先证明一个定理再来分析。

定理 2.1. $\Delta V_{Y=y}$存在上界。

此定理证明需要用到两个引理。

引理 2.1. $E(\log(\varphi(X)))\leq \log(E(\varphi(X)))$成立,其中随机变量$X\geq 0,\varphi(X)\geq 0$。

证明：

由于 log 函数满足 $\log(\lambda x_1+(1-\lambda)x_2)\geq \lambda\log(x_1)+(1-\lambda)\log(x_2),\lambda\in[0,1]$，所以 log 函数为上凸函数。

若令$\lambda=\frac{x_2-x}{x_2-x_1}$，其中$x\in[x_1,x_2]$

则有$\log(x)\geq\frac{x_2-x}{x_2-x_1}\log(x_1)+\frac{x-x_1}{x_2-x_1}\log(x_2)$

不等式等价于$\frac{1}{x-x_1}[\log(x)-\log(x_1)]\geq\frac{1}{x_2-x_1}[\log(x_2)-\log(x_1)]$

令$x\to x_1$，则$(x_2-x_1)\log'(x_1)\ \geq[\log(x_2)-\log(x_1)]$。

再令$x_0=\sum_{i=1}^{m}\lambda_i x_i$，其中$\sum_{i=1}^{m}\lambda_i=1$，$\lambda_i>0$

则对任意 i 都有$\lambda_i(x_i-x_0)\log'(x_0)\ \geq\lambda_i[\log(x_i)-\log(x_0)]$成立

故$\sum_{i=1}^{m}\lambda_i(x_i-x_0)\log'(x_0)\ \geq\sum_{i=1}^{m}\lambda_i[\log(x_i)-\log(x_0)]$

即$\log(\sum_{i=1}^{m}\lambda_i x_i)\ \geq\sum_{i=1}^{m}\lambda_i\log(x_i)$成立。

又由于

$$E\big(\log\big(\varphi(X)\big)\big)=\int\log\big(\varphi(x)\big)\,dF(x)$$
$$=\lim_{n\to\infty}\sum_{k=1}^{\infty}\log\big(\varphi(x_k)\big)\Big(F\Big(\frac{k}{n}\Big)-F\Big(\frac{k-1}{n}\Big)\Big)$$

$$\log\Big(E\big(\varphi(X)\big)\Big)=\log\Big(\int\varphi(x)\,dF(x)\Big)$$

$$= \log\left(\lim_{n\to\infty}\sum_{k=1}^{\infty}\varphi(x_k)\left(F\left(\frac{k}{n}\right) - F\left(\frac{k-1}{n}\right)\right)\right)$$

且 log 函数为连续函数，而且有$\log(\sum_{i=1}^{m}\lambda_i x_i) \geq \sum_{i=1}^{m}\lambda_i \log(x_i)$，故只须令$\lambda_k = F\left(\frac{k}{n}\right) - F(\frac{k-1}{n})$，$m \to \infty$，则$\int \log(\varphi(x))\, dF(x) \leq \log(\int \varphi(x)\, dF(x))$成立，即$E(\log(\varphi(X))) \leq \log(E(\varphi(X)))$。

引理 2.2. 若b^*为对数最优投资组合的权重，则有$E\frac{b^T X}{b^{*T} X} \leq 1$，对其它任意权重 b 成立。

证明：

令$W(b_\lambda, F) = \int \log(b_\lambda^{\boldsymbol{T}}\boldsymbol{X})\, dF(\boldsymbol{X})$, $b_\lambda = \lambda b + (1-\lambda)b^*$，其中 b 为其它解，由于当λ =0 时，$b_0 = b^*$，则根据定义有$W(b_0, F) = W(b^*, F) = \max_{\boldsymbol{b}\in A}\int \log(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X})\, dF(\mathbf{X})$为最大值，即$W(b_0, F) \geq W(b_k, F)$，$k \in [0,1]$，故根据导数的定义知当$\lambda \to 0_+$时，有$\frac{dW(b_\lambda, F)}{d\lambda} \leq 0$成立。

也就是说

$$\begin{aligned}
\lim_{\lambda\to 0_+}\frac{dW(b_\lambda, F)}{d\lambda} &= \lim_{\lambda\to 0_+}\frac{1}{\lambda}[W(b_\lambda, F) - W(b_0, F)] \\
&= \lim_{\lambda\to 0_+}\frac{1}{\lambda}[E(\log(\lambda b^T X + (1-\lambda)b^{*T}X)) - E(\log(b^{*T}X))] \\
&= E\left(\lim_{\lambda\to 0_+}\frac{1}{\lambda}\log\left(\lambda\frac{b^T X}{b^{*T}X} + 1 - \lambda\right)\right) \\
&= E\left(\lim_{\lambda\to 0_+}\frac{1}{\lambda}\log\left(1 + \lambda(\frac{b^T X}{b^{*T}X} - 1)\right)\right) \\
&= E\left(\frac{b^T X}{b^{*T}X} - 1\right) \leq 0
\end{aligned}$$

最后的等式运用$\lim_{x\to 0}\log(1 + c * x)/x = c$ 的结论，这里 c=$(\frac{b^T X}{b^{*T}X} - 1)$，

故$E\frac{b^T X}{b^{*T}X} \leq 1$。

定理 2 证明：

$$\Delta V_{Y=y} = \int \log(\mathbf{b}^{*\boldsymbol{T}}\mathbf{X})\, dF(X|Y=y) - \int \log(\mathbf{b}_{\mathbf{0}}^{\boldsymbol{T}}X)\, dF(X|Y=y)$$

$$= \int \log\frac{\mathbf{b}^{*\boldsymbol{T}}\mathbf{X}}{\mathbf{b}_{\mathbf{0}}^{\boldsymbol{T}}X} dF(X|Y=y)$$

$$=\int \log (\frac{\mathbf{b}^{*T}\mathbf{X}}{\mathbf{b}_0^T \mathrm{X}}\frac{\mathrm{f(x)}}{\mathrm{f}_{x|Y=y}(\mathrm{x})})\mathrm{dF}(\mathrm{X}|\mathrm{Y}=\mathrm{y})+\int \log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dF}(\mathrm{X}|\mathrm{Y}=\mathrm{y})$$

$$=\int \log (\frac{\mathbf{b}^{*T}\mathbf{X}}{\mathbf{b}_0^T \mathrm{X}}\frac{\mathrm{f(x)}}{\mathrm{f}_{x|Y=y}(\mathrm{x})})\mathrm{dF}(\mathrm{X}|\mathrm{Y}=\mathrm{y})+\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$

$$\leq \log\int \frac{\boldsymbol{b}^{*T}\boldsymbol{X}}{\boldsymbol{b}_0^T X}\frac{\mathrm{f(x)}}{\mathrm{f}_{x|Y=y}(\mathrm{x})}\mathrm{dF}(\mathrm{X}|\mathrm{Y}=\mathrm{y})+\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$ (由

引 理 1)

$$=\log\int \mathrm{f}_{x|Y=y}(\mathrm{x})\frac{\mathbf{b}^{*T}\mathbf{X}}{\mathbf{b}_0^T \mathrm{X}}\frac{\mathrm{f(x)}}{\mathrm{f}_{x|Y=y}(\mathrm{x})}\mathrm{dx}+\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$

$$=\log\int \frac{\mathbf{b}^{*T}\mathbf{X}}{\mathbf{b}_0^T \mathrm{X}}\mathrm{dF(X)}+\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$

$$\leq \log 1+\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$ （由引理 2）

$$=\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(\mathrm{x})}{\mathrm{f(x)}}\mathrm{dx}$$

证毕。

在此定理 1 的基础上，由于 Y 信息的引入，资本增长速度增量的期望 $E(\Delta V)=\int \Delta V_{Y=y} dG(y)$，G(y)为 Y 的分布函数。E(ΔV)也存在上界。

$$\mathrm{E}(\Delta \mathrm{V}) \leq \int \Delta V_{Y=y} dG(y)$$

$$=\int\int \mathrm{f}_{x|Y=y}(\mathrm{x})\log\frac{\mathrm{f}_{x|Y=y}(x)}{\mathrm{f}(x)}dx dG(y)$$

$$=\int\int \mathrm{f}_{x|Y=y}(\mathrm{x})\cdot \mathrm{g}(y)\cdot\log\frac{\mathrm{f}_{x|Y=y}(x)\mathrm{g}(y)}{\mathrm{f}(x)\mathrm{g}(y)}dx\, dy$$

$$=\int\int \mathrm{h(x,y)}\log\frac{\mathrm{h(x,y)}}{\mathrm{f}(x)\mathrm{g}(y)}dx\, dy$$

其中h(x, y)为 X, Y 的联合分布。

E(ΔV)存在上界$\int\int \mathrm{h(x,y)}\log\frac{\mathrm{h(x,y)}}{\mathrm{f}(x)\mathrm{g}(y)}dx\, dy$，这在信息论中称作 Y 与 X 的共同信息（mutual information）【1】。与资产收益比的共同信息体现了不同类别的信息的价值。

3 对数最优策略的贪婪与增长最优性质

资产收益服从独立同分布的情形时，最佳的资本增长策略是保持某一固定配置，最佳的动态配置由于这个条件变成了静态了。第 2 节讨论的投资信息的引入使得，独立同分布假设不再符合实际，每一轮的最佳资本增长配置受信息的影响变得不同，是动态同时又是局部最优的。因为资本增长策略优化的指标为$\log\left(\prod_i \boldsymbol{b}_i^T \boldsymbol{X}_i\right) = \sum_i \log\left(\boldsymbol{b}_i^T \boldsymbol{X}_i\right)$。$\{\boldsymbol{b}_i\}$之间的取值并没有相互约束。所以对于每个$\boldsymbol{i}$, 局部最优化$\log\left(\boldsymbol{b}_i^T \boldsymbol{X}_i\right)$是最优化$\log\left(\prod_i \boldsymbol{b}_i^T \boldsymbol{X}_i\right)$的充要条件，即资产增长策略的“贪婪”性质并不违背全局最优的要求。

下面的定理分析了对于对数最优策略与其他非对数最优策略在长期的差别。

定理 3: 令$\mathrm{S}_\mathrm{n}^* = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{b}_\mathbf{i}^{*T} \mathbf{X}_\mathbf{i}$, 其中$\mathrm{b}_\mathrm{i}^*$是对数最优策略的权重, $\mathrm{S}_\mathrm{n} = \prod_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{b}_\mathbf{i}^T \mathbf{X}_\mathbf{i}$是任意其他策略进行配置的资产数量。$\mathrm{S}_\mathrm{n}^*$渐近优于其它投资策略$\mathrm{S}_\mathrm{n}$。

证明：

根据引理 2.2 知，有$\mathrm{E}\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} \le 1$成立

又

$$\Pr(\mathrm{S}_\mathrm{n} > \mathrm{n}^2 \cdot \mathrm{S}_\mathrm{n}^*) = \Pr\left(\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} > \mathrm{n}^2\right) = \int_{\mathrm{n}^2}^{+\infty} \mathrm{dF}\left(\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*}\right)$$

$$\le \frac{1}{\mathrm{n}^2}\int_{\mathrm{n}^2}^{+\infty} \frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} \mathrm{dF}\left(\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*}\right)$$

$$\le \frac{1}{\mathrm{n}^2}\int_{0}^{+\infty} \frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} \mathrm{dF}\left(\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*}\right)$$

$$\le \frac{1}{\mathrm{n}^2}\mathrm{E}\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} \le \frac{1}{\mathrm{n}^2}$$

即,

$$\Pr\left(\frac{1}{\mathrm{n}}\log\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} > \frac{1}{\mathrm{n}}\log\ \mathrm{n}^2\right) \le \frac{1}{\mathrm{n}^2}, \quad \sum_{\mathrm{n}=1}^{\infty} \Pr\left(\frac{1}{\mathrm{n}}\log\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} > \frac{2\mathrm{logn}}{\mathrm{n}}\right) \le \sum_{\mathrm{n}=1}^{\infty}\frac{1}{\mathrm{n}^2} < \infty$$

又

$$\Pr\left(\overline{\lim_{\mathrm{n}\to\infty}}\left\{\frac{1}{\mathrm{n}}\log\frac{\mathrm{S}_\mathrm{n}}{\mathrm{S}_\mathrm{n}^*} > \frac{2\log\ \mathrm{n}}{\mathrm{n}}\right\}\right)$$

$$= \lim_{k\to\infty} \Pr\left(\bigcup_{n=k}^{\infty}\left\{\frac{1}{n}\log\frac{S_n}{S_n^*} > \frac{2\log\ n}{n}\right\}\right)$$

$$\le \lim_{k\to\infty}\sum_{n=k} \Pr\left(\left\{\frac{1}{n}\log\frac{S_n}{S_n^*} > \frac{2\log\ n}{n}\right\}\right)$$

$$= 0$$

这暗示了对股票市场上每个股票序列，存在自然数$N > 0$，对$\forall n > N$，有$\frac{1}{n}\log\frac{S_n}{S_n^*} \le \frac{2\log n}{n}$。因此，我们有

$$\overline{\lim}_{n\to\infty}\frac{1}{n}\log\frac{S_n}{S_n^*} \le 0\text{，以概率 1。}$$

因而，我们有S_n^*渐近优于其它投资策略S_n。

4 理论到实践：真实的概率分布未知

第 2 节中讲到，理论上，信息Y的引入不会降低资本最优增长速度。因为理论假设$F(\mathbf{X}|\mathbf{Y} = y)$是已知的，然而在实际应用中$F(\mathbf{X}|\mathbf{Y} = y)$是利用统计模型进行估计的，估计就会带来误差，这些误差对结果的影响需要进行控制。本节讨论一种对资产配置权重向量进行L_1约束来达到误差控制的方法，通过数学上的推导证明其方法的合理性。

优化指标的化简

设$\mathbf{X}$的期望收益为$\boldsymbol{\mu}$,协方差阵为$\boldsymbol{\Sigma}$，记$R = b^T X$，则$E(R) = E(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{\mu}$。

最精确的对数最优策略为$\boldsymbol{b} = \underset{\boldsymbol{b}\in B}{argmax}\int log(\boldsymbol{b}^T\boldsymbol{X})dF(\boldsymbol{X})$,然而这个优化问题依赖于整个$\boldsymbol{X}$概率密度分布，精确的随机优化需要大量的计算。我们可以通过泰勒展开近似对问题进行合理的化简,使得对优化过程的分析以及计算的实现变得更容易。

$$\int \log(\boldsymbol{b}^T\boldsymbol{X})\, dF(\boldsymbol{X}) = E(\log(\boldsymbol{b}^T\boldsymbol{X}))$$

$$\approx E\left(\log E(\boldsymbol{b}^T\boldsymbol{X}) + \frac{\boldsymbol{b}^T\boldsymbol{X} - E(\boldsymbol{b}^T\boldsymbol{X})}{E(\boldsymbol{b}^T\boldsymbol{X})} - \frac{\left(\boldsymbol{b}^T\boldsymbol{X} - E(\boldsymbol{b}^T\boldsymbol{X})\right)^2}{2\left(E(\boldsymbol{b}^T\boldsymbol{X})\right)^2}\right)$$

$$= E(\log(\boldsymbol{b}^T\boldsymbol{\mu})) + \frac{E(\boldsymbol{b}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{\mu})}{E(\boldsymbol{b}^T\boldsymbol{X})} - \frac{E((\boldsymbol{b}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{\mu})^2)}{2\left(E(\boldsymbol{b}^T\boldsymbol{X})\right)^2}$$

$$= \log(\boldsymbol{b}^T\boldsymbol{\mu}) + 0 - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T]\boldsymbol{b}$$
$$= \log(\boldsymbol{b}^T\boldsymbol{\mu}) - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}$$

令$\mathrm{M}(\mathbf{b},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \log(\boldsymbol{b}^T\boldsymbol{\mu}) - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}$，则我们的问题转化为：

$$\max M(\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\Sigma})$$
$$\mathrm{s.t.}\ \boldsymbol{b} \in B$$

最优投资组合$\boldsymbol{b}^{opt}$的求解

若$\boldsymbol{b} \in B$所表示的约束形式不复杂的话，可以直接求解$\boldsymbol{b}$。本小节中，以$\boldsymbol{B} = \{\mathbf{b}|\boldsymbol{b}^T\boldsymbol{e} = 1，\boldsymbol{b}^T\boldsymbol{\mu} \ge \mathrm{c}\}$为例，分析最优解应该满足的条件。第一个约束是自然的，第二个约束表示投资收益的期望的要求。

由于约束条件中既含有等式约束,又含有不等式约束,因此我们用 Karush-Kuhn-Tucker 条件来对该优化问题进行求解。

记$\mathrm{F}(\mathbf{b},\alpha,\beta) = -\log(\boldsymbol{b}^T\boldsymbol{\mu}) + \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b} + \alpha(\boldsymbol{b}^T\boldsymbol{e} - 1) + \beta(\mathrm{c} - \boldsymbol{b}^T\boldsymbol{\mu})$,

其中$\beta \ge 0$

令$\partial\mathrm{F}(\mathbf{b},\alpha,\beta)/\partial\mathbf{b} = \mathbf{0}$，并与$\beta(c - \mathbf{b}^T\boldsymbol{\mu}) = 0$，$\beta \ge 0$，$\mathbf{b}^T\boldsymbol{e} - 1 = 0$联立，得

$$\begin{cases} -\frac{\boldsymbol{\mu}}{\mathbf{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{e} - \beta\boldsymbol{\mu} = \mathbf{0} & (4.1) \\ \beta(c - \boldsymbol{b}^T\boldsymbol{\mu}) = 0 & (4.2) \\ \beta \ge 0 & (4.3) \\ \boldsymbol{b}^T\boldsymbol{e} - 1 = 0 & (4.4) \end{cases}$$

接下来我们考虑该方程组的求解。

在方程（4.1）两边同时左乘$\boldsymbol{b}^T$，得

$$0 = \boldsymbol{b}^T\left[-\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{e} - \beta\boldsymbol{\mu}\right]$$
$$= -\frac{\boldsymbol{b}^T\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{b}^T\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{b}^T\boldsymbol{e} - \beta\boldsymbol{b}^T\boldsymbol{\mu}$$
$$= -1 - \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\cdot 1 - \beta\boldsymbol{b}^T\boldsymbol{\mu}$$
$$= -1 + \alpha - \beta\boldsymbol{b}^T\boldsymbol{\mu}$$

(1) 若$\beta = 0$，则由上式可得$0 = -1 + \alpha - 0$，即 $\alpha = 1$。

将$\alpha = 1$，$\beta = 0$代入方程（4.1）得，

$$-\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \boldsymbol{e} = \boldsymbol{0} \qquad (4.5)$$

通过方程（4.5），可求出方程最优组合$\widehat{\boldsymbol{b}}^{\mathrm{opt}}$的值。

(2) 若$\beta > 0$，则由方程（4.2）可知，$\boldsymbol{b}^T\boldsymbol{\mu} = \mathrm{c}$。

故有$0 = -1 + \alpha - \beta\mathrm{c}$，即$\alpha = 1 + \beta\mathrm{c}$。

因此，我们可以得到如下方程组：

$$\begin{cases} -\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{e} - \beta\boldsymbol{\mu} = \boldsymbol{0} \\ \boldsymbol{b}^T\boldsymbol{\mu} = \mathrm{c} \\ \alpha = 1 + \beta\mathrm{c} \end{cases} \qquad （4.6）$$

因此，通过求解方程组（4.6），通过求解该方程组，即可求出方程最优组合$\widehat{\boldsymbol{b}}^{\mathrm{opt}}$的值。

由方程（4.5）以及方程组（4.6）可以看出，$\widehat{\boldsymbol{b}}^{\mathrm{opt}}$的求解依赖于参数$\boldsymbol{\mu}$和$\boldsymbol{\Sigma}$的值。若我们对$\boldsymbol{\mu}, \boldsymbol{\Sigma}$的估计出现了偏差，则会导致我们的最优证券投资组合的估计出现偏差，因此我们需要对我们的估计过程进行控制，以保证估计偏差能够得到有效控制。

$\mathbf{L_1}$控制条件下的稳健处理

将不同 b 对应的$\mathrm{M}(\mathbf{b}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 取值称为对应的效用值。设$\widehat{\boldsymbol{b}}^{opt}$是我们对最优化证券组合比重的估计。由于证券投资组合的优劣是通过效用函数值来体现的，所以我们需要准确了解我们的投资组合$\widehat{\boldsymbol{b}}^{opt}$的效用函数值。我们需要通过参数$\boldsymbol{\mu}$和$\boldsymbol{\Sigma}$的估计值，从而求得$\widehat{\boldsymbol{b}}^{opt}$的效用值的估计。因此，$\boldsymbol{\mu}$和$\boldsymbol{\Sigma}$的估计误差会影响到$\widehat{\boldsymbol{b}}^{opt}$的效用值估计的精确性。接下来我们讨论如何在优化框架下添加约束，使得$\widehat{\boldsymbol{b}}^{opt}$的估计效用值与$\widehat{\boldsymbol{b}}^{opt}$的真实效用值的偏差的上界得到控制。

首先，我们需要先对模型做一个假定。

(A1)假定$\mathrm{E}(\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$，即$\hat{\boldsymbol{\mu}}$是$\boldsymbol{\mu}$的无偏估计。

由该假定我们可知，$\mathrm{E}\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right) = \hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}$。由大数定律我们可知，对于$\forall\varepsilon > 0$，当样本量$\mathrm{n} \to \infty$时，有$\mathrm{P}\left(\left|\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right| > \varepsilon\right) \to 0$。所以给定充分小的$\varepsilon > 0$，存在$n_0 \in N_+$，当样本量大于$n_0$时，估计值$\hat{\boldsymbol{\mu}}$满足$\left|\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right| \le \varepsilon$。

另外，我们还需要证明一个引理。

引理 4.1：$\forall p, q, q_1, q_2 \in \mathbb{R}$,满足$q_1 \le q \le q_2$,则有

$$|p - q| \le max\{p - q_1, q_2 - p\}$$

证明：若$p \ge q$，则$p - q_1 \ge p - q \ge 0$，即$p - q_1 > |p - q|$。

若$p < q$，则$q_2 - p \ge q - \mathrm{p} > 0$，即$q_2 - p > |p - q|$。

因此，我们有$|p - q| \le max\{|p - q_1|, |q_2 - p|\}$

接下来我们考察$\hat{\boldsymbol{b}}^{opt}$的估计效用值$\mathrm{M}\left(\hat{\boldsymbol{b}}^{opt}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$与其真实效用值$\mathrm{M}\left(\hat{\boldsymbol{b}}^{opt}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$的偏差。

$$\left|\mathrm{M}\left(\hat{\boldsymbol{b}}^{opt}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right) - \mathrm{M}\left(\hat{\boldsymbol{b}}^{opt}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right|$$

$$= \left|\log(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}) - \frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \log(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}) + \frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}\right|$$

$$= \left|\left(\log(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}) - \log(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu})\right) + \left(\frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt} - \frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt}\right)\right|$$

$$\leq \left|\log(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}) - \log(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu})\right|$$

$$+ \left|\frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt} - \frac{1}{2\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt}\right|$$

$$= \left|\log\left(\frac{\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}}{\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}}\right)\right| + \frac{1}{2}\left|\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}\right|$$

$$\leq \left|\log\left(\frac{\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} + \varepsilon}{\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}}\right)\right| + \frac{1}{2}\left|\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}\right|$$

$$\leq \left|\log\left(1 + \frac{\varepsilon}{\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}}\right)\right| + \frac{1}{2}\left|\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}\right|$$

对于不等式左边第一项，由于$\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} \geq c$，所以

$$\left|\log\left(1 + \frac{\varepsilon}{\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}}\right)\right| \leq \log\left(1 + \frac{\varepsilon}{c}\right)$$

对于不等式右边第二项 ，由引理 4.1，

因为

$$\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} + \varepsilon\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt} < \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt} < \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \varepsilon\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}$$

所以

$$\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}$$

$$\leq \max\left\{\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt},\right.$$

$$\left.\frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \varepsilon\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \varepsilon\right)^2}\hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\hat{\boldsymbol{b}}^{opt}\right\}$$

而由于$-\varepsilon \leq \hat{\boldsymbol{b}}^{opt^T}\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu} \leq \varepsilon$，

则一方面，

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}$$

$$\leq\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\left[\widehat{\boldsymbol{\Sigma}}-\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right]\widehat{\boldsymbol{b}}^{opt}$$

$$\leq\frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\widehat{\boldsymbol{\Sigma}}-\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right)\widehat{\boldsymbol{b}}^{opt}\right)$$

另一方面，

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}$$

$$\leq\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\left[\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right]\widehat{\boldsymbol{b}}^{opt}$$

$$\leq\frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\widehat{\boldsymbol{b}}^{opt}\right)$$

所以

$$\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$\leq\max\left\{\frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\widehat{\boldsymbol{\Sigma}}-\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right)\widehat{\boldsymbol{b}}^{opt}\right),\ \frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\widehat{\boldsymbol{b}}^{opt}\right)\right\}$$

$$\approx\frac{1}{c^2}\left|\widehat{\boldsymbol{b}}^{opt^T}(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}})\widehat{\boldsymbol{b}}^{opt}\right|$$

令$\widehat{\boldsymbol{b}}^{opt}=\left(\hat{b}_1,\dots\dots,\hat{b}_n\right)^T$，$\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}$的第 i 行第 j 列元素为$\sigma_{ij}$。

则

$$\left|\widehat{\boldsymbol{b}}^{\boldsymbol{opt}^T}(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}})\widehat{\boldsymbol{b}}^{\boldsymbol{opt}}\right|=\left|\sum_{i=1}^{n}\hat{b}_i\left(\sum_{j=1}^{n}\hat{b}_j\sigma_{ij}\right)\right|\leq\sum_{i=1}^{n}\left|\hat{b}_i\right|\left|\sum_{j=1}^{n}\hat{b}_j\sigma_{ij}\right|$$

$$\leq \sum_{i=1}^{n}\left|\hat{b}_i\right|\sum_{j=1}^{n}\left|\hat{b}_j\right|\left|\sigma_{ij}\right| \leq \sum_{i=1}^{n}\left|\hat{b}_i\right|\sum_{j=1}^{n}\left|\hat{b}_j\right|\left|\sigma_{ij}\right|$$

$$\leq \sum_{i=1}^{n}\left|\hat{b}_i\right|\left(\sum_{j=1}^{n}\left|\hat{b}_j\right|\sum_{i=1}^{n}\left|\sigma_{ij}\right|\right)$$

$$\leq \sum_{i=1}^{n}\left|\hat{b}_i\right|\max_i\left(\sum_{j=1}^{n}\left|\hat{b}_j\right|\sum_{j=1}^{n}\left|\sigma_{ij}\right|\right)$$

$$= \left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2 \max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$$

故得：

$$\left|\mathrm{M}\left(\widehat{\boldsymbol{b}}^{opt}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right) - \mathrm{M}\left(\widehat{\boldsymbol{b}}^{opt}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right|$$

$$\leq \left|\log\left(1 + \frac{\varepsilon}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right| + \frac{1}{2}\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$\leq \log\left(1 + \frac{\varepsilon}{c}\right) + \frac{1}{2c^2}\left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2 \max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$$

$$\approx \frac{1}{2c^2}\left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2 \max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$$

因此，如果我们选取一个合适大小的常数$c_0 > 0$，并控制$\sum_{i=1}^{n}\left|\hat{b}_i\right| \leq c_0$，则只需要确保样本量足够大，方差的估计与真实值的偏差$\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}$不至于过大（即满足$\max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$较小），就能保证我们对$\widehat{\boldsymbol{b}}^{opt}$的效用值的估计不会偏离真实值过多。

因此，添加额外的$\sum_{i=1}^{n}|b_i| \leq c_0$约束能够达到稳健的优化效果，并且使得最优效用的误差有一个可以控制的上界。

5 金融市场模拟

5.1 模拟目的，假设与数据介绍

模拟目的

本节使用真实的数据对前文理论进行检验。开发一套策略，仅通过历史的价格数据，寻找价格模式，以此进行连续的资产选择，达到资本的增长。

模拟假设

1. 简化交易费用的影响，模拟中涉及以日为单位的资产平衡与再选择，现实世界中这样的操作对应着未知的交易费用。我们采用【2】论文中所建立的线性交易费用模型对影响进行近似。
2. 流动性。虽然模拟中所考虑的资产流动性都非常大，但现实中进行无约束的平衡再选择还是有一些困难的，因此模拟中不考虑这种约束。

数据介绍

模拟实验所用数据是纽约证券交易所(NYSE)上 36 支股票的数据，时间段为 1962 年 6 月 3 号到 1984 年 12 月 31 号，共计 5651 个交易日，该数据在【3,4】论文中均有使用。

5.2 模型

1.基于市场相似度的信息提取

本次模拟只使用股价数据，通过历史数据中出现的相似价格模式对未来的价格进行预测。记$\mathrm{x(i,k)}$为第 i 支股票在第 k 天的价格比。X(k)表示第 k 个交易日的价格比数据。$\mathrm{X(k:l)} = \begin{bmatrix} x_{1k} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{pk} & \cdots & x_{pl} \end{bmatrix}$为第 k 至 l 天所有的股票价格比数据,可以看作是第 k+1 天前 l-k+1 天的市场背景环境。

对于任意一个交易日 t，固定它市场背景环境长度，都可以在历史中寻找具有类似市场背景环境的交易日。这些交易日的数据对于 t+1 天的价格是具有预测

价值的。

定义市场背景相似度

$$\mathrm{Similar}\left(X\left(k：\mathrm{l}\right), X(k+m: l+m)\right)=\mathrm{corr}^{*}\left(X(k:\mathrm{l}), X(k+m: l+m)\right)$$

其中corr^{*}表示将$\mathrm{X}\left(k：\mathrm{l}\right), X(k+m: l+m)$两个矩阵列向拉直成两个向量之后求皮尔森相关系数的运算。

寻找市场背景相似交易日

通过确定市场背景长度n以及相似度阀值ρ这两个参数，定义与第 k 个交易日市场背景相似的交易日的日期指标集为：

$$\mathrm{C}(k, n, \rho)=\left\{\mathrm{n}<i<k \mid Similar\left(X\left(i-n：\mathrm{i}-1\right), X(k-n: k-1)\right)>\rho\right\}$$

2.配置优化模型

在第 k+1 个交易日前进行资产选择前，寻找到相似的交易日集合，相似交易日的收益比的经验分布即作为第 k+1 个交易日的收益比估计。值得提醒的是，相似交易日的选取收到市场长度n以及系数阀值ρ的影响。最适应市场的（n,ρ）不能事先知道，股票数据的时序性又使得交叉验证不适用此问题。通过整个训练集寻找最佳的（n,ρ）又容易引起过拟合的问题。

为了解决上述问题，也使得资产选择更加稳健。考虑多个（n,ρ）的组合，每个组合看成是一个“专家”，最终决策综合考虑了专家的意见。本文中专家的选取方法如下，取定一个 N 作为n的上界，考虑$2 \leq \mathrm{n} \leq \mathrm{N}$的所有 n。取定一个 P 作为ρ的上界,在$[0, \mathrm{P}]$范围内固定间距取 10 个点作为专家ρ的取值。参数 N,P 的选取对模型结果的影响比（n,ρ）更稳健，这点在灵敏性分析部分能够得到展示。

对于单个专家(n,ρ)，第 k+1 天的配置优化模型为：

$$\mathbf{b}_{\mathrm{k}+1}(n, \rho)=\underset{\boldsymbol{b} \in A}{\operatorname{argmax}} \prod_{i \in \mathrm{C}(k, n, \rho)}\left(\boldsymbol{b} \cdot \boldsymbol{X}_{\boldsymbol{i}}\right)$$

$b \in A$为资产配置向量所要满足的约束条件。根据第 4 节我们添加 L1 范数不大于 1 的约束，以最小化估计误差的影响上界。现实中这等价于不卖空的约束。

通过如下公式，自适应的动态加权综合专家意见：

$$\mathbf{b}_{\mathrm{k+1}} = \frac{s(n,\rho)\mathbf{b}_{\mathrm{k+1}}(n,\rho)}{\sum_{n,\rho} s(n,\rho)\mathbf{b}_{\mathrm{k+1}}(n,\rho)}$$

$s(n,\rho)$表示以$\mathbf{b}_{\mathrm{k+1}}(n,\rho)$配置在$\mathrm{C}(k,n,\rho)$集合里的交易日能够获得的总收益。

3.交易费用影响

交易费用是考察策略的重要环节，通过仿真模拟环节简单的规则可以探究交易费用的影响，但最准确的交易费用一定在真实市场上才能计算出来。

本文采取【2】文献中提出的线性交易费用模型。第 k+1 天配置的交易费用为：

$$\alpha \cdot \|b_{k+1} - b_k\|_1$$

在考虑收益时，需要减去相应比例的费用。α为调整系数，取值 0 代表无交易费用，依次增大表示交易费用增大。本文考虑不同α下，策略的表现。

5.3 模拟结果

总财富

将参数设成 n=10, $\rho = 0, \alpha = 0$，上述策略在 NYSE 数据实施后，得到的累积资本的增长情况如图 1。

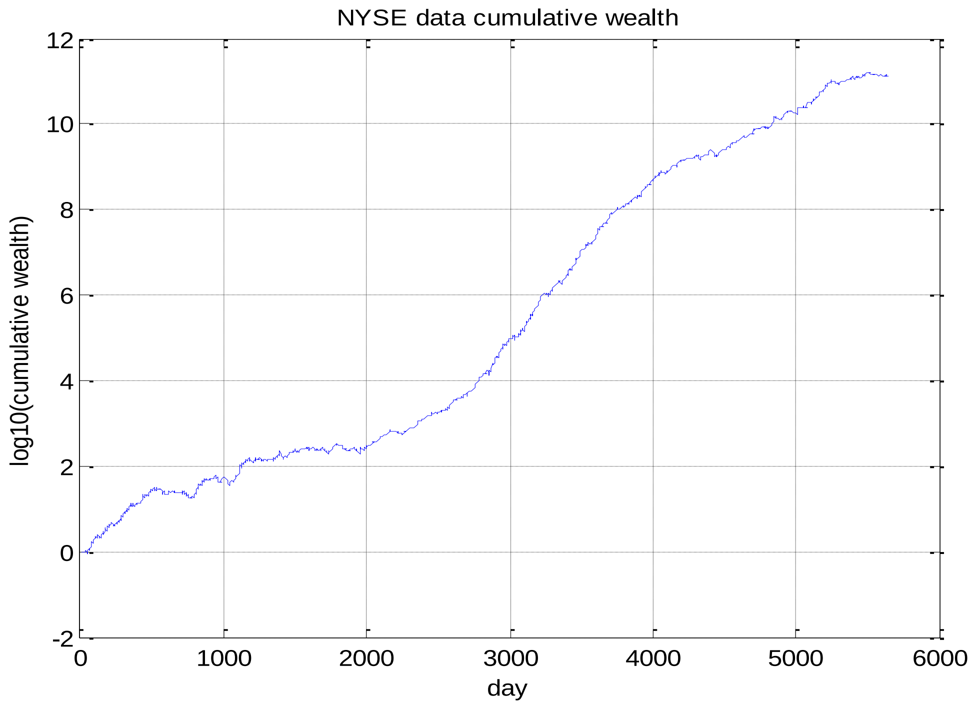

图 1.NYSE 数据累积资本

结合资本累积图可以看出：经过 5651 个交易日，即 5651 次资产配置后，不考虑每次调整所需的交易费用，资产增加的倍数可以达到10^{11}这个数量级。而且，资产在每个阶段都能够稳步的增长，不会出现数量级上的大幅度降低。这样良好表现的原因是：大数定律保证了在一段时间内策略中优化的增长速度期望的稳定。并且基于市场背景相似度的预测符合市场规律，能够提高增长速度的期望。（还是要把交易费考虑进去！）

交易费用影响

实际应用中，需要考虑交易费用的影响。结合 5.2 节中线性交易费用模型，考虑每次资产调整需扣除调整头寸α比例的资金。α由 0 增加至 0.01，代表交易扣费程度增加，对应的累积资产如图 2 所示。

通过图 2，可以看出线性费用影响下，累积资产的对数与费用系数α服从一个线性的关系。在$\alpha = 0.0045$这个临界点处，资产数量对数是 0，即对应着不增不减。当α小于此临界点时，累积的资产数量能够随着时间以指数增长。

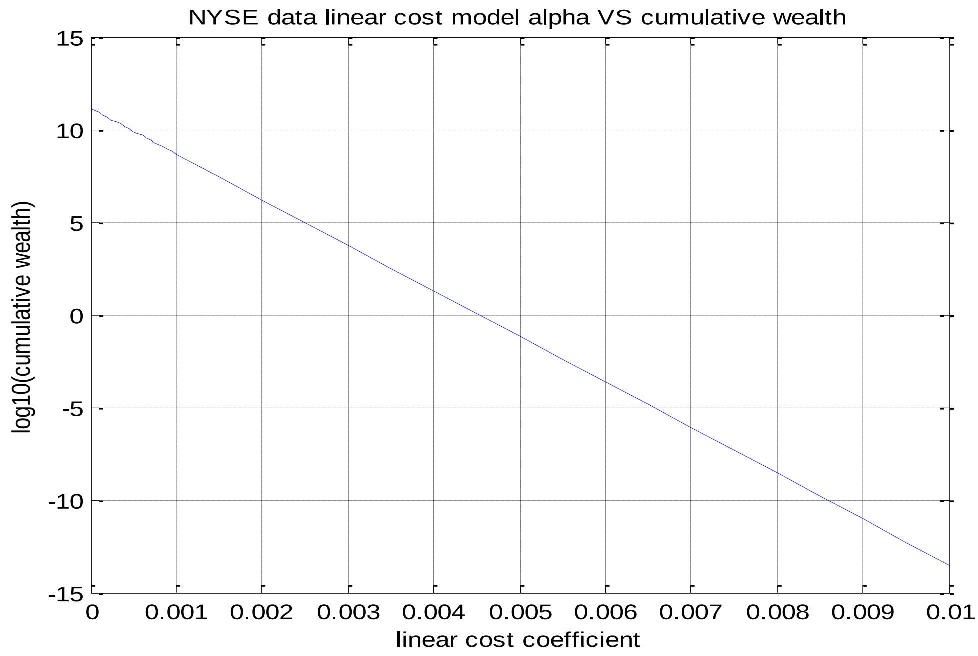

图 2.交易费用影响分析

参数灵敏性分析

前文所诉的方法设定了最大市场背景长度为 10。为了分析这一参数的不同设定对结果是否有很灵敏的影响，将最大市场背景长度调整为 0 至 25 的整数，得到对应的最终累积资本数量，如图 3 所示。

图 3 中结果显示，最大市场背景长度参数在 0-5 之间时，参数越大，结果越好。在 5-25 之间，结果的数量级保持稳定。在实际应用中，可以通过训练集选出最佳的市场背景长度，由于这个参数的稳健性，它在样本外数据的表现也可以得到保证，这可以解决数据挖掘的过拟合的问题。

同样，对参数ρ进行灵敏性分析。将 n 取 10，ρ分别取{-0.15，-0.05，0，0.05，0.1，0.15}，得到最后不同累积资本，如图 4 所示。从图中可以看出，ρ取 0 时，能够达到最好的效果。ρ 为负时，提取出的相似交易日会有反作用，非最优是正常的。ρ 太大时，符合要求的交易日太少，效果也不会最佳。

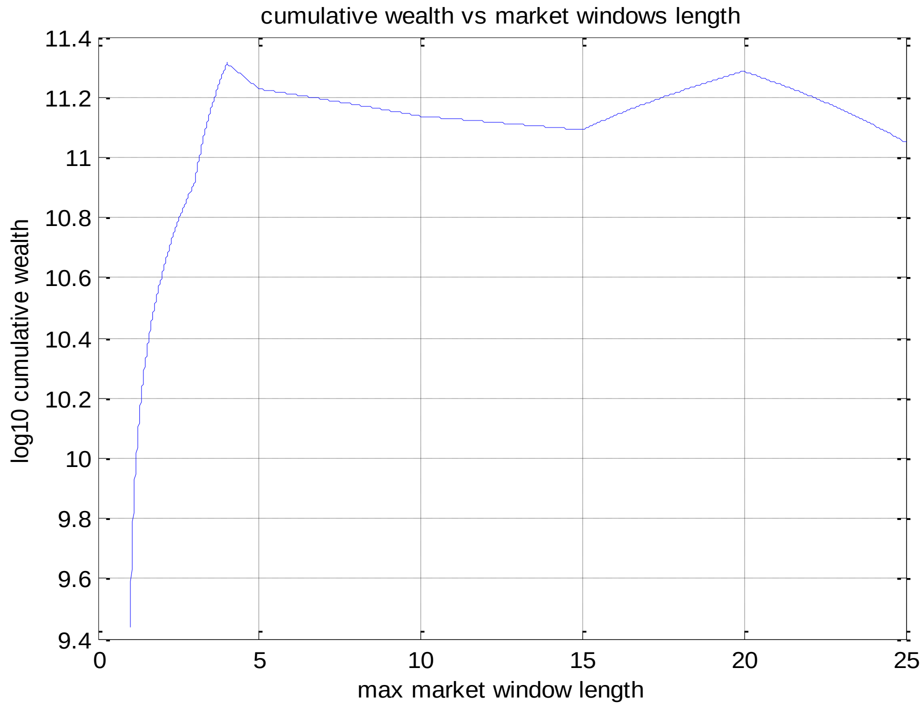

图 3.最大市场长度灵敏性分析

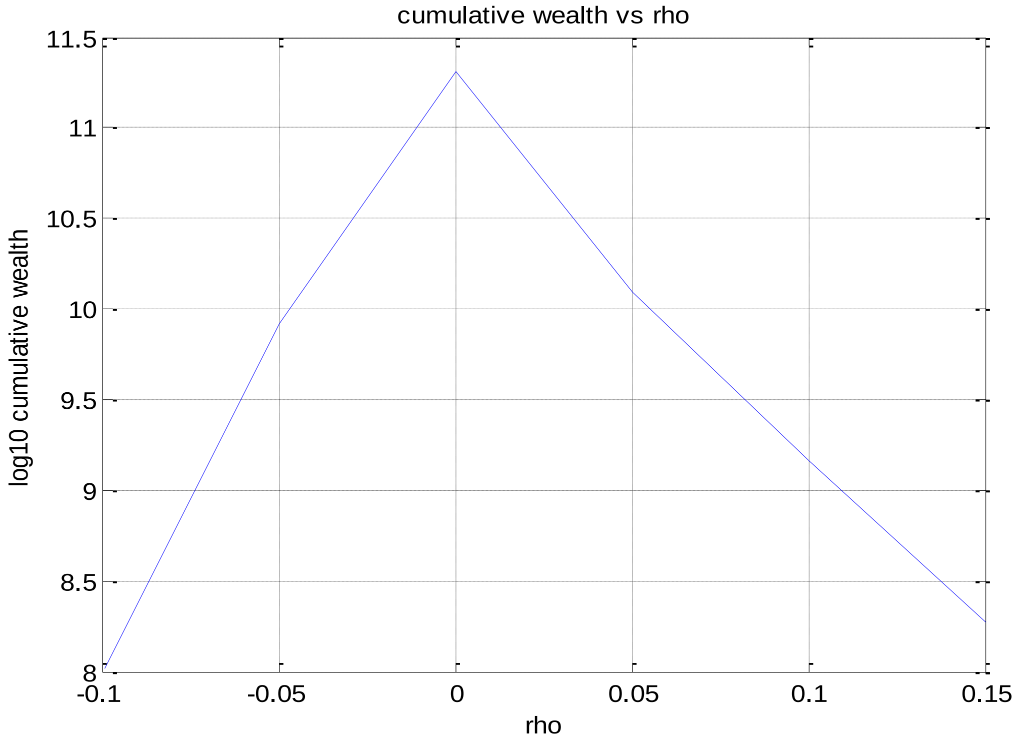

图 4.相关性阈值灵敏性分析

6.结论

本文由经典赌博资金管理问题引入，得到对数最优策略的定义。在此基础上，本文分析了信息对资本增长速度增量的影响，并通过理论推导证明了对数最优策略的渐近最优性。此外，本文考虑了L_1约束下均值和方差误差对资本收益误差的影响，从而证明了在此框架下对数最优策略对资产配置具有较好的稳健性。NYSE 股票数据的模拟实证研究表明对数最优策略具有较高的收益率和优异的稳定性。

7.参考文献

【1】Cover T M, Thomas J A. Elements of information theory[M]. John Wiley & Sons, 2012.

【2】Blum A, Kalai A. Universal portfolios with and without transaction costs[J]. Machine Learning, 1999, 35(3): 193-205.

【3】Cover T M. Universal portfolios[J]. Mathematical finance, 1991, 1(1): 1-29.

【4】Singer Y. Switching portfolios[J]. International Journal of Neural Systems, 1997, 8(04): 445-455.